

Cause-Effect Dynamics of Computer and Network Systems for QoS

Paper ID: 967

**Nong Ye, Steve Yau, Dazhi Huang, Mustafa Baydogan, Billibaldo Martinez Aranda, and
Autawut Roontiva**

**School of Computing, Informatics, and Decision Systems Engineering
Arizona State University
Tempe, AZ 87287, USA**

**Patrick Hurley
Air Force Research Laboratory
Rome, NY 13441, USA**

Abstract

To provide Quality of Service (QoS) demanded by many online services (e.g., e-commerce), computer and network systems need to have QoS monitoring and adaptation which must be based on cause-effect dynamics relations among service activities, the state of system resources, and the QoS performance of service processes. This paper presents our study on cause-effect dynamics models for one of computer and network services, the voice communication service with the throughput of network data as the QoS feature of interest. Experiments are conducted to obtain computer and network data under various service conditions that are set up using three service parameters: the sampling rate, the number of clients and the buffer size. The experimental data is then analyzed. We uncover four major types of cause-effect dynamics. We found a set of five system state variables concerning the memory, CPU, process and IP resources that are significantly affected by the service parameters and are closely linked to the network throughput performance of the voice communication service. New insights are also gained about satisfying high QoS demands by properly increasing the size of the buffer which holds the network data before the data transmission.

Keywords

Quality of service (QoS), cause-effect relations, voice communication service, statistical analysis and modeling

1. Introduction

Quality of Service (QoS) has been essential for conventional, off-line services, and has become increasingly important as many conventional services are moved online with computer and network systems providing services. In [1] we present QoS requirements of various network applications for online services (e.g., web browsing, email, file transfer, audio and video broadcasting, audio and video on demand, audio and video conferencing, voice over IP, etc.) with QoS metrics on timeliness (e.g., response time, delay and jitter), precision (e.g., bandwidth and loss rate) and accuracy of services (e.g., error rate).

When competing service requests with specific QoS requirements come to a computer network system providing services, the system must determine if its limited system resources can accommodate the service requests and provide the services at the required QoS level. The system also needs to determine what service configuration, resource configuration and service-resource binding should be used to achieve the required QoS level [2, 3]. On a computer and network system, competing service requests and resulting service activities change the state of limited system resources which in turn affect QoS provided to users by the system [2, 4]. Such dynamic relations of service activities, system state and QoS performance provide the basis for determining the satisfaction of QoS requirements

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 2010		2. REPORT TYPE		3. DATES COVERED 00-00-2010 to 00-00-2010	
4. TITLE AND SUBTITLE Cause-Effect Dynamics of Computer and Network Systems for QoS				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Arizona State University ,School of Computing, Informatics, and Decision Systems Engineering,Tempe,AZ,87287				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES U.S. Government or Federal Rights License					
14. ABSTRACT To provide Quality of Service (QoS) demanded by many online services (e.g., e-commerce), computer and network systems need to have QoS monitoring and adaptation which must be based on cause-effect dynamics relations among service activities, the state of system resources, and the QoS performance of service processes. This paper presents our study on cause-effect dynamics models for one of computer and network services, the voice communication service with the throughput of network data as the QoS feature of interest. Experiments are conducted to obtain computer and network data under various service conditions that are set up using three service parameters: the sampling rate, the number of clients and the buffer size. The experimental data is then analyzed. We uncover four major types of cause-effect dynamics. We found a set of five system state variables concerning the memory, CPU, process and IP resources that are significantly affected by the service parameters and are closely linked to the network throughput performance of the voice communication service. New insights are also gained about satisfying high QoS demands by properly increasing the size of the buffer which holds the network data before the data transmission.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 6	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

and actions of QoS configuration and adaptation.

However, models of such activity-state-QoS relations are not readily available from the design of computer and network systems which provides mostly algorithm-based operational models. Activity-state-performance dynamic models from existing studies focus mostly on specific resources (e.g., router, CPU, memory, and hard disk) individually and limited aspects of dynamic models. Few models of activity-state-QoS dynamics during realistic operations of computer and network systems exist at a more comprehensive, system scale to account for a wide range of hardware and software resources (including CPU, memory, physical disk, caches, buffers, network interface, IP, TCP, UDP, terminal service, etc.), their interactions, their state changes with service activities, and their effects on QoS performance. Service and system configuration for QoS satisfaction should not simply consider the service effect on an individual resource or change the configuration of an individual resource because all system resources interact with and place constraints on one another. The QoS performance depends on activity-state-QoS dynamics at the system scale, i.e., effects of service activities on all system resources and QoS of all competing service processes/threads.

Due to the lack of models for activity-state-QoS dynamics at a more comprehensive, system scale, existing studies on QoS configuration and adaptation often bypass the issue of establishing models of activity-state-QoS dynamic relations. Those studies focus only on the evaluation of QoS according to user-defined weights of QoS attributes without getting into models of realistic activity-state-QoS dynamics. Hence, those studies do not account for how the performance level of various QoS attributes change dynamically with competing, dynamic service demands and the varying state, constraints and interactions of limited system resources.

The lack of models for realistic activity-state-QoS dynamic relations at the system scale produces a significant gap in bringing existing work on QoS configuration and adaptation to real-world applications. Our studies aim at establishing models of activity-state-QoS dynamics models at the system scale to fill in this gap. Using an empirical approach, we collect system dynamics data of service activities, resource state and QoS performance with services running on a real computer and network system under various service conditions and use experimental data to characterize and build activity-state-QoS models. Our empirical studies involve various service types, e.g., voice communication and motion detection, since activity-state-QoS dynamics may vary with different computer and network services. In this paper, we present our methodology of data collection, data analysis and data modeling, and illustrate our methodology using the voice communication service as a case study.

In Section 2, we describe the method and facility of data collection which are illustrated through the description of the experiment to collect system dynamics data under various service conditions and system configurations for a voice communication service. In Section 3, we present the methodology to analyze the experimental data, uncover activity-state-QoS relationships, and build activity-state-QoS models. In section 4, we discuss the analytical and modeling results. Section 5 concludes the paper.

2. Methodology of Data Collection and Experiment for Voice Communication Service

Our empirical studies focus on computers with an Windows operating system because they are widely used. Windows operating systems provide Windows performance objects [5] which cover various aspects of process activities, resource state and service performance for many objects of resources and processes. Some examples of performance objects are Physical Disk, Memory, System, Process, Processor, IP, UDP, and TCP. Each object has a number of counters which reflect various aspects of activity, state and performance of the object. For example, the Memory object has a counter, *Available Bytes*, shows one aspect of the memory state. The IP object has a counter, *Fragmented Datagrams/sec*, which reflects the data transmission performance at the IP level.

System dynamics data of service activities, resource state and QoS performance must be collected under various levels of service activities and system configurations to reveal varying activity-state-QoS dynamics. Our experiment for the voice communication service involves two parameters of service activities, the sampling rate of recording voice data and the number of clients sending service requests, and one parameter of system configurations, the size of the buffer on a voice communication server for storing the voice data before transmitting the data to a client over the network. In the experiment, the voice communication service is set up in an online radio broadcasting context in which multiple clients simultaneously request and receive the voice communication service from a server which sends out real-time voice data streams of various quality levels controlled by the data sampling rate. Multiple clients run on their own computers with only one client on one computer. The computer network system for the experiment

stands alone without connections with any other computer and network system to avoid interferences.

The voice communication service is a communication-intensive application. The QoS attribute for the voice communication service concerns mainly the throughput of voice data transmission. Hence, we want to select parameters of service activities and system configurations that are expected to affect the QoS performance of the voice data communication service concerning the throughput of voice data transmission from the server to the clients. Two parameters of service activities are selected for the experiment: 1) the sampling rate (Sa) which is used to record the voice data stream and thus determine the quality of voice data, and 2) the number of clients (C). The parameter of system configuration, the size of the buffer (B) on the server for storing the voice data before transmitting the data to a client over the network, is selected because the buffer size directly affects the throughput of voice data transmission. As shown in Table 1, each parameter has five levels in the experiment such that we collect data with sufficient granularity for data analysis to obtain ASQ relationships and models. Hence, we have 125 (5 x 5 x 5) experimental conditions. The five levels of the sampling rate are denoted by Sa1, Sa2, Sa3, Sa4, and Sa5. The five levels of the number of clients are denoted by C1, C2, C3, C4, and C5. The five levels of the buffer size are denoted by B1, B2, B3, B4, and B5.

Table 1: Service and system parameters and their levels in the experiment

Parameter	Level 1	Level 2	Level 3	Level 4	Level 5
Sampling rate (Sa)	44,100 Hz	88,200 Hz	132,300 Hz	176,400 Hz	220,500 Hz
Number of clients (C)	1	2	3	4	5
Buffer size (B)	16 Kbytes	24 Kbytes	32 Kbytes	40 Kbytes	48 Kbytes

ASQ models represent cause-effect relations among service activities (A), state of resources (S), and QoS performance (Q) of service processes. The two parameters of the sampling rate and the number of clients directly drive service activities and are considered as an A variable in the ASQ models. The parameter of the buffer size affects the state of the buffer. With complex interactions of the buffer with other system resources during the process of the voice communication service, the parameter of the buffer size is also expected to affect the state of other system resources. Hence, the parameter of the buffer size is also considered as an A variable that is expected to affect the state of system resources and thus QoS performance. System dynamics data of resource state and QoS performance of the voice communication service for S and Q variables in ASQ models are collected using eight Windows performance objects, including Physical Disk, Memory, System, Process, IP, UDP, TCP, and Server. These objects are selected because they are expected to be involved in the voice communication service.

The experimental run under each of the 125 experimental conditions includes one minute of the voice communication service for a given level of the three service parameters. Sixty data observations under each experimental condition are collected with the sampling rate of 1 observation per second. The data is recorded in log files. Experimental data is collected on the server since the server data reflects the effect of multiple clients requesting the service. The analysis of the data is performed on the server data to obtain ASQ relations and models.

3. Methodology of Data Analysis and Modeling

The data from the experiment with the full set of 125 service conditions is used to uncover the relations of the service activity parameters with the resource state variables and QoS performance variables collected from the Windows performance object utility. The following statistical data analyses are carried out.

- 1) A-SQ relation discovery and categorization. For each state or QoS variable, the Analysis of Variance (ANOVA) in Statistica7 is performed with the three activity parameters of the sampling rate, the number of clients, and the buffer size (A's) as the independent variables and each state or QoS variable (S or Q) as the dependent variable to determine the main and interaction effects of A's on the S or Q variable. Each independent variable has five levels. ANOVA determines if the main and interaction effects of the three independent variables on the dependent variable are statistically significant. If a main or interaction effect is statistically significant based on the ANOVA results, the Tukey's honest significant difference (HSD) test in Statistica7 is performed to determine how different levels of one or more activity parameters affect the state or QoS variable. The qualitative A-S or A-Q relation of the activity parameters (A) with a state variable (S) or an QoS variable (Q) is revealed from the Tukey's HSD test results and is further categorized into a certain type of A-SQ relations.
- 2) Development of the ASQ relation map. ANOVA and Tukey's HSD test results reveal the A-SQ relations.

Among the S variables that appear in the A-SQ relations, we determine which of the S variables have direct cause-effect relations with a given Q variable through an inference based on the design and operations of computers and networks. The A-S relations and the S-Q relations are then captured in an ASQ relation map by representing an A, S, or Q variable as a node and a relation as a directed link between nodes.

- 3) ASQ modeling. For each S or Q variable in the ASQ relation map, the qualitative A-S or S-Q relation represented in the map is refined into a quantitative prediction model of the S variable from one or more A variables or the Q variable from the related S variables. The Multivariate Adaptive Regression Splines (MARS) technique for nonlinear regression models [6] using the *earth* package in the *R* software (<http://cran.r-project.org/web/packages/earth/earth.pdf>) is used to build a regression model for an A-S or S-Q relationship.

4. Results and Discussions

ANOVA for each of the state and QoS variables reveals 28 state and QoS variables which have at least one significant main or interaction effect (with the p -value less than 0.05) of the three activity parameters. For each significant main or interaction effect on each state or QoS variable, the Tukey's HSD test is performed to determine how different levels of one or more activity parameters affect the state or QoS variable. Using the Tukey's test results, the 28 state and QoS variables, which have at least one significant main or interaction effect based on the ANOVA results, are grouped into the following five categories of the A-SQ relations. Table 2 lists the variables in each category of the A-SQ relations. The variables in category 5 are likely affected by not only the voice communication service but also system routine activities which together produce the inconsistent change pattern of these variables with the service parameters of the voice communication service along with sophisticated interactions. Hence, the variables in category 5 should not be considered as accurate measures of system state and QoS performance that are directly or solely linked to the voice communication service. In summary, the 21 state and QoS variables in Categories 1-4 are directly related to the voice communication service, and show four major categories of the A-SQ relations.

Table 2: Five categories of A-SQ relations and the state and QoS variables in each category

Category	State and QoS Variables	
1. Increase with Sa and C and B	% Committed Bytes In Use_Memory	Committed Bytes_Memory
2. Increase with Sa and C, decrease with B	% Privileged Time_Process % User Time_Process Datagrams/sec_UDP Datagrams Sent/sec_UDP File Control Operations/sec_System Fragmented Datagrams/sec_IP IO Other Operations/sec_Process	% Processor Time_Process Context Switches/sec_System Datagrams/sec_IP Datagrams Sent/sec_IP File Control Bytes/sec_System Fragments Created/sec_IP IO Other Bytes/sec_Process
3. Increase with C, stable with Sa except at one end, inverse-U change with B	Thread Count_Process	Page Faults/sec_Memory
4. Decrease with Sa, C and B	Available Bytes_Memory Available MBytes_Memory	Available KBytes_Memory
5. Inconsistent change with Sa, C and B and sometimes strong interaction of Sa, C and B	% Registry Quota In Use_System Datagrams Received Delivered/sec_IP Datagrams Received/sec_IP Page Faults/sec_Process	Avg. Disk sec/Transfer_Physical Disk Processor Queue Length_System Datagrams Received/sec_UDP

Among the 21 state and QoS variables, some variables present similar information. While summarizing the ASQ relations into an ASQ relation map with each node representing a variable and a link representing a direct relationship between two variables, we can keep only one variable among a group of variables that present similar information. The following five state variables:

- *Committed Bytes_Memory*
 - *% Processor Time_Process*
 - *IO Other Operations/sec_Process*
 - *Thread Count_Process*
 - *Page Faults/sec_Memory*,
- and the following QoS variable:
- *Fragments Created/sec_IP*
- are kept in the ASQ relation map (shown in Figure 1) and used for ASQ modeling.

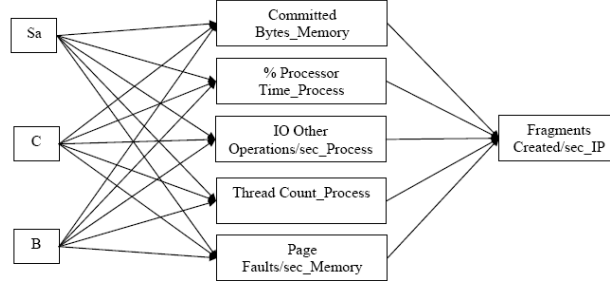


Figure 1: The ASQ relation map

For each state variable in the ASQ relation map, we use the MARS technique to build a regression model to represent the quantitative A-S relation of the three service parameters with the state variable. For the QoS variable, we build a regression model to represent the quantitative S-Q relation of the five state variables with the QoS variable. We denote the A variables, the sampling rate, the number of clients and the buffer size, by X_S , X_C , and X_B , respectively. The state variables, *Committed Bytes_Memory*, *% Processor Time_process*, *IO Other operations/sec_Process*, *Thread Count_Process*, and *Page Faults/sec_Memory*, are represented by S_{CB} , S_{PT} , S_{IO} , S_{TC} , and S_{PF} , respectively. The QoS variable, *Fragments Created/sec_IP*, is denoted by Q_{FC} . Table 3 summarizes the results of the MARS regression models for AS and SQ relationships. The R-square values in Table 3 show that the AS and SQ models fit the data well. The MARS model for *Committed Bytes_Memory* is shown below as an example of AS models.

$$\begin{aligned}
 S_{CB} = & 498357996.5 + 240.35 * \max(0, X_S - 132300) - 282.33 * \max(0, 132300 - X_S) \\
 & + 14828753.24 * \max(0, X_C - 2) - 14541707.34 * \max(0, 2 - X_C) - 86 * \max(0, 40960 - X_B) \\
 & + 101.32 * \max(0, 132300 - X_S) * \max(0, X_C - 4) - 15.48 * \max(0, 132300 - X_S) * \max(0, 4 - X_C) \\
 & + 0.0055 * \max(0, 88200 - X_S) * \max(0, 40960 - X_B) + 71.06 * \max(0, 2 - X_C) * \max(0, 40960 - X_B) \\
 & + 0.0033 * \max(0, 88200 - X_S) * \max(0, X_C - 2) * \max(0, 40960 - X_B) \\
 & + 0.0098 * \max(0, 132300 - X_S) * \max(0, X_C - 4) * \max(0, X_B - 24576) \\
 & + 0.0153 * \max(0, 132300 - X_S) * \max(0, X_C - 4) * \max(0, 24576 - X_B) \\
 & + 0.01 * \max(0, 132300 - X_S) * \max(0, X_C - 4) * \max(0, X_B - 32768) \\
 & + 0.0007 * \max(0, 132300 - X_S) * \max(0, 4 - X_C) * \max(0, 40960 - X_B)
 \end{aligned}$$

Table 3: A summary of MARS regression models for AS and SQ relations

S or Q Variable	Number of S or Q Variables in the Model	R-Square Value
Committed Bytes_Memory	3	0.9987
% Processor Time_Process	3	0.9889
IO Other Operations/sec_Process	3	0.9840
Thread Count_Process	3	0.9921
Page Faults/sec_Memory	3	0.8160
Fragments Created/sec_IP	3	0.9921

5. Conclusions

This paper presents our methodologies of data collection, data analysis and data modeling to establish activity-state-QoS models for enabling QoS configuration and adaptation. We illustrate the methodology using the voice communication service as a case study. In this case study, We uncover a number of state and QoS performance variables that are significantly affected by the three service activity parameters of the sampling rate, the number of clients and the buffer size, including *Committed Bytes_Memory*, *% Processor Time_Process*, *IO Other Operations/sec_Process*, *Thread Count_Process*, *Page Faults/sec_Memory*, *Fragments Created/sec_IP*, and others. We also reveal four major categories of the ASQ relations. The ASQ relations and the regression models defining the quantitative ASQ relationships will be useful in predicting the change of QoS performance when initially configuring and later adapting the service activity parameters, the resource capacity and the service-resource binding to meet the QoS requirements. Although this case study is based on a communication-intensive voice communication service, the experimental and analytical methodologies are applicable to investigating and developing dynamics models of service activity, system state and QoS cause-effect dynamics for other computer and network services.

Acknowledgements

This work is sponsored in part by the National Science Foundation (NSF) under grant number CCF-0725340 and in part by the Air Force Research Laboratory (AFRL) under award number FA8750-08-2-0155. The U.S. government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of, NSF, AFRL, or the U.S. Government. We would like to thank Professor Hessam Sarjoughian for his constructive comments throughout this research.

References

1. Chen, Y., Farley, T., and Ye, N., 2004, "QoS requirements of network applications on the internet," *Information, Knowledge, Systems Management*, 4(1), 55-76.
2. Ye, N., 2008, *Secure Computer and Network Systems: Modeling, Analysis and Design*, Wiley Publishing, London, UK.
3. Yau, S. S., Huang, D., Zhu, L., and Cai, K.-Y., 2007, "A software cybernetics approach to deploying and scheduling," *FTDCS '07: Proceedings of the 11th IEEE International Workshop on Future Trends of Distributed Computing Systems*, 149-156.
4. Ye, N., 2002, "QoS-centric stateful resource management in information systems," *Information Systems Frontiers*, 4(2), 149-160.
5. Microsoft, 2003, *Window Server 2003 Performance Counters Reference*. Retrieved 03/01, 2009, from <http://technet2.microsoft.com.ezproxy1.lib.asu.edu/windowsserver/en/library/3fb01419-b1ab-4f52-a9f8-09d5eb9ef21033.mspx>
6. Friedman, J. H., 1991, "Multivariate adaptive regression splines," *The Annals of Statistics*, 19(1), 1-67.
7. Mann, H. B., and Whitney, D. R., 1947, "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Mathematical Statistics*, 18(1), 50-60.